

Robust Estimation of the Magnitude Squared Coherence based on Kernel Signal Processing

Ferran De Cabrera, *Student Member, IEEE*, Jaume Riba, *Senior Member, IEEE*, and Gregori Vázquez, *Senior Member, IEEE*
Signal Theory and Communications Department, Technical University of Catalonia (UPC)
{ferran.de.cabrera, jaume.riba, gregori.vazquez}@upc.edu

2017 51st Asilomar Conference on Signals, Systems, and Computers, Date of Conference: 29 Oct.-1 Nov. 2017
DOI: 10.1109/ACSSC.2017.8335477

Abstract—A new outlier-robust approach to estimate the magnitude squared coherence of a random vector sequence, a common task required in a variety of estimation and detection problems, is proposed. The proposed estimator is based on Rényi's entropy, an information theoretic kernel-based measure that proves to be inversely proportional to the determinant of a regularized version of the covariance matrix in the proper Gaussian case. The trade-off between accuracy and robustness in terms of bias and variance is analytically and numerically characterized, showing a dependence on the relative kernel bandwidth and the available data size.

I. INTRODUCTION

Robust estimation of the covariance matrix is an important task involved in a wide range of signal processing applications (see [1] and references therein). It is well-known that the sample covariance matrix coincides with the maximum likelihood estimator (MLE) in the case of independent and identically distributed vector sequences. However, in many applications the samples follow either unknown distributions or non-stationary anomalies that violate the distributional assumptions. When this happens, the goal is to develop estimation methods capable of trading-off some efficiency at the nominal model to gain resistance against the effects of deviations [2]. In this paper, the goal is to estimate the Magnitude Squared Coherence (MSC), a statistic widely used for non-parametric detection of a common signal on two noisy channels [3].

Consider a i.i.d. vector sequence of the form $\mathbf{x}_i = [x_{1i}, x_{2i}]^T$ with covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_1 & \rho\sqrt{\Sigma_1\Sigma_2} \\ \rho^*\sqrt{\Sigma_1\Sigma_2} & \Sigma_2 \end{pmatrix} \quad (1)$$

where ρ is the coherence factor or Pearson coefficient. In most signal processing applications we are interested on estimating the Magnitude Squared Coherence (MSC) defined as

$$c = |\rho|^2. \quad (2)$$

The MSC is a fundamental statistics involved in the Locally Most Powerful Test (LMPIT) ([4] & [5]) for deciding whether or not two random sequences are correlated. On the other hand, the Shannon mutual information between two Gaussian random variables is a monotonically increasing function of the MSC given by $-\log(1 - c)$, where $1 - c$ is just the Hadamard Ratio, i.e. the determinant of the covariance matrix over the product of its diagonal elements.

Among the solutions proposed to solve this problem, the Tyler's iterative estimator of the covariance matrix (see seminal papers [6] & [7]), which yields robustness by assuming that the data is drawn from heavy-tailed distributions, occupies a prominent place in the robust estimation literature. However, this estimator requires a known mean of the data observed or a prior and robust estimate of it.

The objective of this paper is estimating c from N samples of \mathbf{x}_i ($i = 1, \dots, N$) in a robust manner against the presence of outliers in the measurements. By using kernel signal processing we are able to relate an entropy measure to the determinant of the covariance matrix, while taking advantage of the property that entropy measures depend on the probability of anomalous events, instead of their magnitude, and are insensitive to the mean. This property succeeds on moving the interest from the typical heavy-tail Gaussian assumption of the data and focusing on large-valued impulsive outlier model. The derived estimator will be analyzed and compared to Tyler's performance.

II. ESTIMATION OF MULTIVARIATE INFORMATION POTENTIAL

The Information Potential (IP), the argument of the log in the Rényi Entropy [8], is defined as

$$V = \int f^2(\mathbf{x}) d\mathbf{x} \quad (3)$$

where $f(\mathbf{x})$ is the multivariate density function of the data with $\mathbf{x} \in \mathbb{C}^M$. We will explore the fact that, for $f(\mathbf{x})$ being the p.d.f. of a proper Gaussian distribution (nominal conditions), the determinant of the covariance matrix is inversely proportional to the IP:

$$V = \frac{1}{(2\pi)^M |\Sigma|} \quad (4)$$

From the previous authors paper [9] and following a similar rationale, we can obtain an estimate of the IP based on Gaussian kernels, which yields

$$\hat{V} = \frac{1}{N^2} \sum_{1 \leq i \leq N} \sum_{1 \leq j \leq N} k_{\mathbf{W}}(\mathbf{x}_i - \mathbf{x}_j) = \frac{1}{N} + \frac{\hat{U}}{(2\pi)^M |\mathbf{W}|} \quad (5)$$

with \hat{U} an unbiased estimator of the IP with the following form:

$$\hat{U} = \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} k_{\mathbf{W}}(\mathbf{x}_i - \mathbf{x}_j) \quad (6)$$

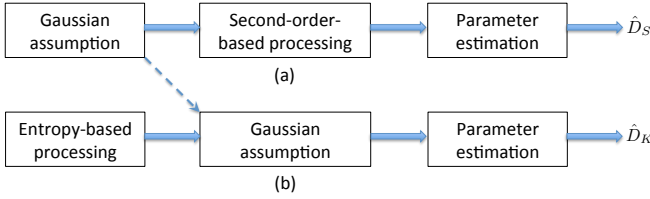


Fig. 1. Main rationale to achieve robustness: compared to (a), the Gaussian assumption is postponed after a prior entropy-based processing in (b).

being $k_{\mathbf{W}}(\mathbf{z}) = e^{-\mathbf{z}^H \mathbf{W}^{-1} \mathbf{z}}$ a Gaussian kernel function with positive definite bandwidth matrix $\mathbf{W} \in \mathbb{R}^{M \times M}$, and the mean estimator \bar{U} , called Modified Information potential (MIP), is defined as

$$\bar{U} = E[\hat{U}] = \frac{|\mathbf{W}|}{|\mathbf{W} + 2\boldsymbol{\Sigma}|}. \quad (7)$$

On the basis of this background, we can state that the procedure will be based on estimating first the MIP and then inferring the matrix covariance from it in a second step, assuming nominal conditions. Hence, given the relation on Eq. (4), we can build an estimate of the MSC based on an estimate of the MIP. The purpose is to construct an estimator of the form

$$\hat{D}_K = g_{\mathbf{W}}(\hat{U}) \quad (8)$$

where $g_{\mathbf{W}}(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ is a kernel-dependent monotonic decreasing function. The main rationale behind this approach is sketched in Fig 1. While estimator \hat{D}_S is derived by first making a Gaussian assumption on the available data and then estimating the desired parameter, our proposed estimator is instead derived by first estimating a non-parametric, entropy-based measure of the data using kernel methods, and then relating the obtained biased (by the kernel) estimate to the desired parameter under the Gaussian assumption. As will be shown, this delayed Gaussian assumption yields a good compromise between near optimality and robustness.

III. GENERALIZED COHERENCE

This rationale coincides with the previous work [9] from the authors about the robust estimation of the covariance determinant on the univariate case. However, this paper extends the idea of robust estimate for the bivariate case given its interest on studying the relation between two signals. It is worth mentioning that in the multivariate case the MSC can be extended with the Generalized Coherence (GC) [10], which is defined as a function of the Hadamard ratio with the following form

$$\gamma = 1 - \frac{\det \left(\begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{1,2} \\ \boldsymbol{\Sigma}_{1,2}^H & \boldsymbol{\Sigma}_2 \end{bmatrix} \right)}{\det(\boldsymbol{\Sigma}_1) \det(\boldsymbol{\Sigma}_2)} \quad (9)$$

with $\boldsymbol{\Sigma}_k = E[(\mathbf{x}^{(k)} - \bar{\mathbf{x}}^{(k)})(\mathbf{x}^{(k)} - \bar{\mathbf{x}}^{(k)})^H]$ ($k = 1, 2$), and $\boldsymbol{\Sigma}_{1,2} = E[(\mathbf{x}^{(1)} - \bar{\mathbf{x}}^{(1)})(\mathbf{x}^{(2)} - \bar{\mathbf{x}}^{(2)})^H]$ the covariance and cross-covariance matrices, respectively, with $\mathbf{x}^{(1)} \in \mathbb{C}^{M_1}$, $\mathbf{x}^{(2)} \in \mathbb{C}^{M_2}$ and $\bar{\mathbf{x}}^{(1)}, \bar{\mathbf{x}}^{(2)}$ their means.

Additionally, we can relate the GC with a Canonical Correlation Analysis (CCA) ([11], [12]) problem by rearranging it in the following way:

$$\begin{aligned} \gamma &= 1 - \frac{\det(\boldsymbol{\Sigma}_2) \det(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_{1,2} \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\Sigma}_{2,1})}{\det(\boldsymbol{\Sigma}_1) \det(\boldsymbol{\Sigma}_2)} \\ &= 1 - \det(\boldsymbol{\Sigma}_1^{-1} (\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_{1,2} \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\Sigma}_{2,1})) \\ &= 1 - \det(\mathbf{I} - \mathbf{C}) \end{aligned} \quad (10)$$

with $\mathbf{C} = \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_{1,2} \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\Sigma}_{2,1}$ being the squared coherence matrix ([13], [4]). If we denote λ_i as the eigenvalues of \mathbf{C} , which correspond to the canonical variables of the CCA, we can express the GC as follows:

$$\gamma = 1 - \prod_{i=1}^{\min(M_1, M_2)} (1 - \lambda_i) \quad (11)$$

Eq. (9) allows us to build an estimator of the GC for the multivariate case as a function of the determinants of three covariance matrices, block composite covariance matrix and marginal ones. The idea is then to use a similar procedure as in Eq. 7 to estimate indirectly the desired determinants in order to obtain the GC.

IV. ENTROPY-BASED MAGNITUDE SQUARED COHERENCE ESTIMATION

Consider the bivariate case $M_1 = M_2 = 1$ and $M = 2$ with the i.i.d observed sequences x_{1i} and x_{2i} ($i = 1, \dots, N$) with marginal variances Σ_1 and Σ_2 respectively and covariance matrix determinant $|\boldsymbol{\Sigma}|$ from Eq. (1). The MSC is then defined as

$$c = 1 - \frac{|\boldsymbol{\Sigma}|}{\Sigma_1 \Sigma_2}. \quad (12)$$

For general purpose, we will assume that the marginal variances are not necessarily equal and the kernel bandwidth \mathbf{W} is a diagonal matrix with elements $[\mathbf{W}]_{m,m} = W_m$ and $m = 1, 2$. Then, we define the relative kernel bandwidth \mathbf{w} as a diagonal matrix with elements $[\mathbf{w}]_{m,m} = W_m / \Sigma_m$. For simplicity, we will assume that $[\mathbf{w}]_{1,1} = [\mathbf{w}]_{2,2} = w$ such as $W_1/W_2 = \Sigma_1/\Sigma_2$. If the marginal variances of the original processes are known, only the kernel bandwidth is needed to be estimated, for instance using the iterative method proposed in [9] by adding the previous condition. Otherwise, both kernel bandwidth and marginal variances need to be estimated by the univariate robust estimate of the covariance matrix described in the same work. This relative kernel bandwidth restriction allows us to express the estimate of the MSC \hat{c} as a composite estimator, generally described by w instead of the marginal variances.

From Eq. (12) and (7) the following monotonic relationship between the MIP and the MSC is obtained:

$$\bar{U} = \frac{w^2}{(w+2)^2 - 4c} \quad (13)$$

As a consequence, we are able to obtain a composite estimator \hat{c} of the MSC by means of the method of moments, resulting in

$$\hat{c} = \left(1 - \frac{1}{\bar{U}}\right) \frac{w^2}{4} + w + 1 \quad (14)$$

Considering the final expression of the estimator, we will provide an analysis of bias and variance under nominal conditions, paying special attention to the interplay between N and w and estimator efficiency, and we will confirm its behavior by a numerical analysis.

A. Bias and Variance

By first analyzing the bias by means of Jensen's inequality $E[\hat{c}] - c \geq (1 - \bar{U}^{-1}) \frac{\omega^2}{4} + \omega + 1$, we obtain that the bias of the estimator is strictly negative, but consistent in the sense that \hat{c} tends to c in probability.

The variance of \hat{c} can be analyzed by a small perturbation analysis, as can be seen in the Appendix VII-A, and has the following form

$$\sigma_{\hat{c}}^2 \approx \left(\frac{(w+2)^2 - 4c}{2w} \right)^4 (1 + 3\bar{U}^{-2} \sigma_{\bar{U}}^2) \sigma_{\bar{U}}^2 \quad (15)$$

where the variance of the MIP, $\sigma_{\bar{U}}^2$, is given by the Appendix VII-B from [9] and has the following form:

$$\sigma_{\bar{U}}^2 = E[\hat{U}^2] - \bar{U}^2 = \frac{aN(N-1)(N-2) + bN(N-1)/2}{(N(N-1)/2)^2} \quad (16)$$

with constants a and b as follows:

$$a = w^4 \left(\frac{1}{((w+1)^2 - c)((w+3)^2 - 9c)} - \frac{1}{((w+2)^2 - 4c)^2} \right) \quad (17)$$

$$b = w^4 \left(\frac{1}{w^2((w+4)^2 - 16c)} - \frac{1}{((w+2)^2 - 4c)^2} \right). \quad (18)$$

From Eq. (16), the asymptotic variance for $N \rightarrow \infty$ will then be:

$$N\sigma_{\hat{c}}^2 \rightarrow \frac{1}{4} \left((w+2)^2 - 4c \right)^4 \times \left(\frac{1}{((w+1)^2 - c)((w+3)^2 - 9c)} - \frac{1}{((w+2)^2 - 4c)^2} \right). \quad (19)$$

It can be seen that, for a fixed N , the variance increases without limit as $w \rightarrow 0$ as well for $w \rightarrow \infty$. On the other hand, the MSC variance depends on the own MSC parameter c . In particular, to get insights, it is easily seen that the asymptotic variance reaches a minimum for $w = 0$. Using Eq. (19) we obtain

$$N\sigma_{\hat{c}}^2, w=0 \rightarrow \frac{28}{9} (1-c)^2. \quad (20)$$

This value becomes maximum for $c = 0$ and tends to zero for $c \rightarrow 1$. These issues will be confirmed later on with computer simulations.

Additionally, we do also obtain the following condition for the adequate relative kernel bandwidth design by considering N sufficiently large for b to have no impact on the variance:

$$w_{min} = \sqrt{\frac{6(1-c)}{N}} \leq \sqrt{\frac{6}{N}}. \quad (21)$$

For the problem of estimating the MSC, the minimum kernel bandwidth is inversely proportional to the square root of the data size, N , in contrast with the faster decay of $O(N^{-1})$ given in [9]. It is also worth noting that this result applies only for the proper complex case, and that the obtained decay would instead be $O(N^{-1})$ and $O(N^{-2})$ for the MSC and the variance, respectively, in the bivariate real case, which would make less critical the selection of the bandwidth.

B. Robustness

Finally, let us focus on the robustness. In particular, in order to provide maximum insights on estimating the MSC, we consider a bivariate model of replacement outliers (see [14] and [2]) in which the outliers do not modify the marginal variances:

$$\mathbf{x}_{\varepsilon i} = (1 - z_i) \mathbf{x}_i + z_i \mathbf{y}_i \quad (22)$$

where z_i is defined as a zero-one process with $P(z_i = 1) = \varepsilon$. Consider, for instance, that \mathbf{x}_i and \mathbf{y}_i are i.i.d. and distributed as $\mathcal{CN}(\mathbf{m}, \Sigma_\rho)$ and $\mathcal{CN}(\mathbf{m}, \Sigma_{-\rho})$, respectively, with $\Sigma_\rho = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and $\rho = \sqrt{c}$. In this manner, a contamination rate of $\varepsilon = 0.5$ causes the MSC of $\mathbf{x}_{\varepsilon i}$, easily given by $(1 - 2\varepsilon)^2 c$, to shrink for any c . In contrast, the information-theoretic measure based on the IP will be sensitive to the p.d.f. of the contaminated data, which becomes a Gaussian mixture of two bivariate p.d.f.s. [15]

$$f_\varepsilon(\mathbf{x}) = \frac{1}{\pi(1-c)} \left((1-\varepsilon) e^{-(\mathbf{x}-\mathbf{m})^H \Sigma_\rho^{-1} (\mathbf{x}-\mathbf{m})} + \varepsilon e^{-(\mathbf{x}-\mathbf{m})^H \Sigma_{-\rho}^{-1} (\mathbf{x}-\mathbf{m})} \right). \quad (23)$$

This causes the IP to decrease from its uncontaminated value of $(2\pi)^{-2} (1-c)^{-1}$ such that, based on Eq. (4), the inferred covariance determinant values become inflated. As a result (see Appendix V-B for details) the inferred MSC becomes:

$$c_\varepsilon = 1 - \frac{1-c}{1-2c\varepsilon(1-\varepsilon)} \geq (1-2\varepsilon)^2 c \quad (24)$$

where the right-hand side characterizes the non-robust sample MSC case. The above inequality proves that the entropy-based estimator is more robust than the sample MSC at least for infinitely small ($w \rightarrow 0$) kernel bandwidths, which, according to Eq. (21), requires that $N \rightarrow \infty$. The case of finite N will be analyzed later on with computer simulations to confirm that the robust behavior explained above is still significantly maintained.

V. NUMERICAL RESULTS

In this section we will evaluate the performance of the robust MSC estimate and it will be compared to non-robust estimators, e.g. the sample covariance matrix, as well as Tyler's estimate. The figures were obtained through Monte Carlo simulations and it was assumed that the marginal variances were equal $\Sigma_1 = \Sigma_2$.

Fig. 2 shows the normalized variance of the MSC estimator as a function of w for increasing values of N and two values of the true MSC, analytical (Eqs. (15), (19) and (20)) and

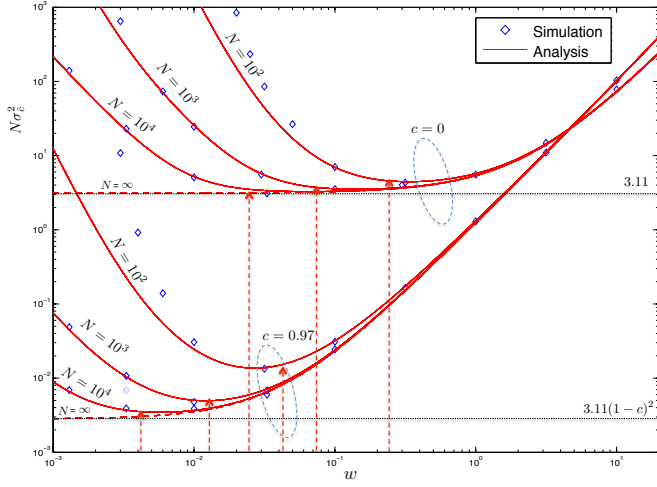


Fig. 2. Normalized variance of the MSC as a function of the relative kernel bandwidth w for different values of the data size, N , and values of the true MSC of $c = 0$ and $c = 0.97$.

numerical. The asymptotic values provided in Eq. (20) are also indicated, which confirm the tendency of the curves for increasing N and decreasing w .

Fig. 3 shows the negative bias of the kernel-based MSC estimate in comparison with the sample covariance (non-robust) and Tyler (robust, with 10 iterations) [6] methods. While the bias of the MSC estimate based on the sample covariance matrix increases linearly with c with a slope proportional to ε , the entropy-based estimate is shown to be much less affected by contamination, and it shows a much more robust behavior than Tyler's approach in the case of moderate values of the MSC. Note also that, as explained in [1], the Tyler's approach requires a prior estimation of the mean, which leads to severe problems in practice depending on the nature of the outlier process.

For illustration, the analytical entropy-based bias for $N = \infty$ is also shown (dashed curves) in order to appreciate the effect of a finite data size on the bias. As explained, the higher is N , the smaller the kernel bandwidth can be fixed according to Eq. (21), which means that the MIP estimate approaches a scaled and shifted version of a true information-theoretical measure (information potential).

VI. CONCLUSION

In this paper we have derived a robust estimate of the MSC in a bivariate case from an estimator of the Information Potential. We have seen that under an outlier hypothesis, a critical issue associated to the sample covariance estimator, the estimate is affected by the probability of the outliers and not by the magnitude of them. The proposed approach provides a solid alternative to other robust approaches, obtaining a better performance in some circumstances. The next step would be to generalize to the multivariate case for $M > 2$, which corresponds to the generalized coherence.

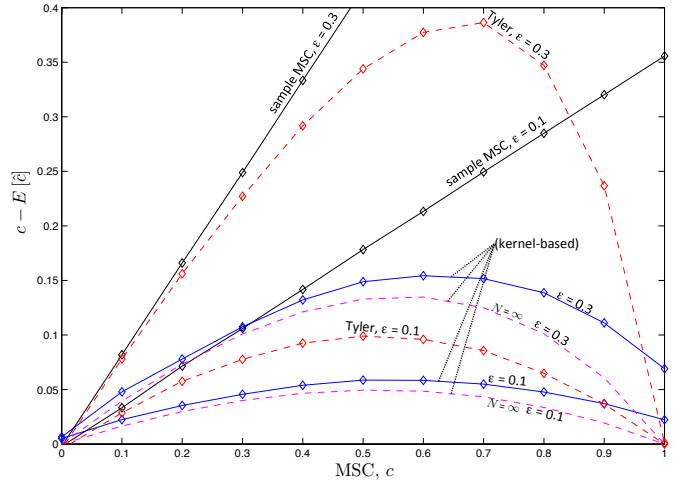


Fig. 3. Negative bias (simulation) of the MSC estimate using several methods, as a function of the true MSC for different values of contamination rate (ε) and data size $N = 500$.

VII. APPENDIX

A. Small perturbation analysis

Defining the zero-mean random variable $dU = \hat{U} - \bar{U}$ with variance σ_U^2 , we can define the estimator \hat{c} as

$$\hat{c} = \left(1 - \frac{1}{\bar{U} + dU}\right) \frac{w^2}{4} + w + 1 \quad (25)$$

By using a second-degree Taylor expansion, and considering $|dU| \ll \bar{U}$, we get

$$\begin{aligned} \hat{c} &\approx ((1 - \bar{U}^{-1}) + \bar{U}^{-2}dU - \bar{U}^{-3}dU^2) \frac{w^2}{4} + w + 1 \\ &\approx \hat{c} + dc \end{aligned} \quad (26)$$

with

$$dc = (\bar{U}^{-2}dU - \bar{U}^{-3}dU^2) \frac{w^2}{4} \quad (27)$$

The variance is then approximated by $\sigma_{\hat{c}}^2 \approx E[(dc)^2]$.

We will make a further approximation by assuming dc is normal, and so we have that $E[(dc)^3] = 0$ and $E[(dc)^4] = 3(\sigma_{\hat{c}}^2)^2$. Finally we have

$$\begin{aligned} \sigma_{\hat{c}}^2 &\approx (\bar{U}^{-4}\sigma_U^2 - 3\bar{U}^{-6}\sigma_U^4) \frac{w^4}{16} \\ &= \left(\frac{(w+2)^2 - 4c}{w^2}\right)^4 \frac{w^4}{16}\sigma_U^2 + \left(\frac{(w+2)^2 - 4c}{w^2}\right)^6 \frac{3w^4}{16}\sigma_U^4 \\ &= \left(\frac{(w+2)^2 - 4c}{2w}\right)^4 (1 + 3\bar{U}^{-2}\sigma_U^2) \sigma_U^2 \end{aligned} \quad (28)$$

which corresponds to Eq. (15).

B. Impact of contamination on IP

From Eq. (23) and using Eq. (4),

$$V_{\varepsilon} = \int f_{\varepsilon}^2(\mathbf{x}) d\mathbf{x} = \frac{(1 - \varepsilon)^2 + \varepsilon}{(2\pi)^2(1 - c)} + 2(1 - \varepsilon)\varepsilon P \quad (29)$$

with

$$\begin{aligned}
 P &= \frac{1}{\pi^4(1-c)^2} \int e^{-\mathbf{x}^H(\Sigma_\rho^{-1} + \Sigma_{-\rho}^{-1})\mathbf{x}} d\mathbf{x} \\
 &= \frac{1}{\pi^4(1-c)^2} \int e^{-\frac{\mathbf{x}^H \mathbf{x}}{(1-c)/2}} d\mathbf{x} \\
 &= \frac{(2\pi)^2 |\mathbf{I}_2(1-c)/2|}{\pi^4(1-c)^2} \left(\int \frac{1}{(2\pi)^2 |\mathbf{I}_2(1-c)/2|} e^{-\frac{\mathbf{x}^H \mathbf{x}}{(1-c)/2}} d\mathbf{x} \right) \\
 &= \frac{1}{(2\pi)^2}. \tag{30}
 \end{aligned}$$

Substituting and rearranging:

$$V_\varepsilon = V_0 (1 - 2c\varepsilon(1 - \varepsilon)). \tag{31}$$

Finally, as the term $(1 - 2c\varepsilon(1 - \varepsilon))$ contaminates the IP in a multiplicative manner, its inverse contaminates the determinant $(1 - c)$ in view of Eq. (4), which finally yields the result shown in Eq. (24) as we wanted to proof.

REFERENCES

- [1] Y. Sun, P. Babu, and D. P. Palomar, "Regularized robust estimation of mean and covariance matrix under heavy-tailed distributions," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3096–3109, June 2015.
- [2] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Process. Magazine*, vol. 29, no. 4, pp. 61–80, 2012.
- [3] D. Cochran, H. Gish, and D. Sinno, "A geometric approach to multiple-channel signal detection," *IEEE Trans. Signal Process.*, vol. 43, no. 9, pp. 2049–2057, 1995.
- [4] D. Ramírez, J. Vía, I. Santamaría, and L. Scharf, "Detection of spatially correlated Gaussian time series," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5006–5015, 2010.
- [5] D. Ramírez, J. Vía, I. Santamaría, and L. Scharf, "Locally most powerful invariant tests for correlation and sphericity of Gaussian vectors," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2128–2141, Apr. 2013.
- [6] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *Ann. Statistic.*, vol. 15, no. 1, pp. 234–251, 3 1987.
- [7] J. T. Kent and D. E. Tyler, "Maximum likelihood estimation for the wrapped Cauchy distribution," *J. Appl. Statist.*, vol. 15, no. 2, pp. 247–254, 1988.
- [8] J. C. Principe, *Information Theoretic Learning: Rényi's Entropy and Kernel Perspectives*. New York: Springer, 2010.
- [9] F. D. Cabrera, J. Riba, and G. Vazquez, "Entropy-based covariance determinant estimation," in *18th IEEE International Workshop on Signal Processing Advances in Wireless Communications, SPAWC 2017*. IEEE, November 2017.
- [10] H. Gish and D. Cochran, "Generalized coherence (signal detection)," *International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 1988.
- [11] J. Vía, I. Santamaría, and J. Pérez, "Canonical correlation analysis (CCA) algorithms for multiple data sets: Application to blind simo equalization," *IEEE 13th European signal processing conference*, 2005.
- [12] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, 1936.
- [13] Y. Wang, L. Scharf, I. Santamaría, and H. Wang, "Canonical correlation for target detection in a passive radar network," *IEEE 50th Asilomar Conference on Signals, Systems and Computers*, 2016.
- [14] R. Maronna, R. Martin, and V. Yohai, *Robust Statistics: Theory and Methods*. New York, NY, USA: Wiley, 2006.
- [15] J. Q. Li and A. R. Barron, "Mixture density estimation," in *Advances in neural information processing systems*, 2000, pp. 279–285.